Available online at www.sciencedirect.com

ScienceDirect

Procedia Technology 11 (2013) 540 – 547

Procedia Technology

The 4th International Conference on Electrical Engineering and Informatics (ICEEI 2013)

A New Algorithm to Estimate the Similarity between the Intentions of the Cyber Crimes for Network Forensics

Mohammad Rasmi*, Aman Jantan

School of Computer Sciences, Universiti Sains Malaysia, Penang 11800 USM, Malaysia

Abstract

Given the rapidly increasing amount of digital crimes, network forensics plays a significant role in the process of analyzing evidence. Current cyber crime investigation techniques are very costly and time consuming because much effort is needed to analyze the overwhelming amount of evidence in each case. More information is required to understand and analyze the factors involved in cyber crime, such as the intention of the crime. This paper proposes a new algorithm called the Similarity of Attack Intention (SAI), which uses cosine similarity as a distance-based similarity measure to estimate similar cyber crime intentions. The algorithm is based on the Attack Intention Analysis (AIA) algorithm to predict new cyber crime intentions and assigned the probability values for these intentions. A similarity metric for the new cyber crimes intentions with others is generated in order to identify the similar intentions. The evaluation of example cyber crime results shows that our proposed algorithm provides better solutions and increases the possibility value of evidences.

Selection and peer-review under responsibility of the Faculty of Information Science & Technology, Universiti Kebangsaan Malaysia.

Keywords: cyber crime; network forensics;attack intention; Attack Intention Analysis algorithm

1. Introduction

The number of digital crimes nowadays is increasing. Consequently, the network forensics plays an important role in the cyber crime analysis process. For example, big companies like Sony group and Google have been penetrated by a sophisticated attack by some computer hackers called themselves "Anonymous" for over a month in

* Corresponding author. Tel.: +6017-4417084
E-mail address: mr77mr@hotmail.com

Selection and peer-review under responsibility of the Faculty of Information Science & Technology, Universiti Kebangsaan Malaysia.
doi:10.1016/j.protcy.2013.12.226

2011. In most of these cases, it takes a lot of time to discover a real crime maker, and until this moment, some of them are still registered as unknown. In reality, according to 2011 Cyber Security Watch Survey [1] 21% of the digital crime events has been caused by "unknown". This indicates that the current network forensics investigation process is time consuming, expensive, and an error-prone process is highly expected in apprehending the real perpetrators.

Palmer [2] defined network forensics as: "The use of scientifically proven techniques to collect, fuse, identify, examine, correlate, analyze, and document digital evidence from multiple, actively processing and transmitting digital sources for the purpose of uncovering facts related to the planned intent, or measured success of unauthorized activities meant to disrupt, corrupt, and or compromise system components as well as providing information to assist in response to or recovery from these activities". The ultimate goal of the network forensics is to provide sufficient evidence to prosecute the perpetrator of the crime as described by [2, 3].

Recently, there are many limitations and specific research gaps related to network forensics [4-6]. Cyber crime analysis is considered as a major challenge for many people who are working in network forensics [5]. Therefore, during the investigation phase, it becomes more difficult to discover the attack and apprehend the perpetrator.

Cyber crime analysis, especially attack intentions, supports investigators in bringing a close criminal cases with greater accuracy in advance to decide the suitable incident response as mentioned in [7, 8]. In addition, analyzing attack intentions is a necessity to produce clear evidence to accelerate the decision processes required for apprehending the real perpetrator.

This paper is proposing a new algorithm to estimate the similarity of the cyber crime intentions with others for network forensics. This algorithm describes the similarity of attack intentions process as mentioned in [13]. The algorithm based on Attack Intentions Analysis (AIA) algorithm [7] to identify the intentions for the new cyber crime. The proposed algorithm is described in section 3, where the process model of a similarity between the intentions of the cyber crimes will be illustrated. The next section will present a related work of network forensics analysis and cyber crimes intentions. Finally, we introduce an experimental example of a cyber crime by using samples of probabilistic values of detection accuracy for the cyber crimes intentions to evaluate our proposed algorithm.

2. Related works

In reality, there are two main network security perspectives, 1. Prevention: such as firewalls and Intrusion Prevention System (IPS). 2. Detection: such as Intrusion Detection System (IDS) [2, 4, 5]. Accordingly, network forensics is considered an extension of network security as well as computer forensics, the last deals with laws and guiding principles of a judicial system. In fact, the network forensics uses scientific techniques to collect, examine, analyze, and document digital evidences from digital sources and network security perspectives to uncover facts related to the cyber crimes. Analyzing digital evidences of the cyber crimes is a actually considered a major challenge in network forensics [5].

In general, the analysis phase in network forensics approaches faces many challenges such as reconstruction methods of attack behavior. Normally we have to go through a full capture of malicious behavior in order to understand the intention of the attack. Also, the classification process and clustering of network events as mentioned by [4], may arise when the protocols' complexity exist. Furthermore, reconstruction methods, which are used to understand the intention of the attack complicate the cyber crime analysis.

Cyber crime analysis has emerged as a comprehensive review of intention, as mentioned in [9], where intruder states are combined with system states to create an intention list. Also, it is possible to estimate threats to predict future attacks depending on their founding intentions. According to Peng et al. [10], attack intentions are realized when it is able to identify the goal of this attack, which presents its path. However, a graph algorithm with methods for intrusive intention recognition used to analyze the attack path in advance to discover the goal of the cyber crimes. Even for an expert, it is difficult to find a method of intrusion [6], which makes the prediction of the attack goals more complicated. Hence, the attack intention as well as the attack analysis is still the main challenge in network forensics [4, 9, 11]. Attack intention analysis, as mentioned in [8] could produce clear evidences to help investigators to make the right decision.

Similarity measurements could improve the quality of the results of the cyber crime evidence analysis. These results minimize the duration time and processing cost of the made decision in the investigation phase of network forensics [12]. In general, most of the similarity methods are based on either distance, or feature, or probabilistic measurements. However, these methods are often used in the alert correlation techniques, which depends on the similarity of the attack attributes [11, 12].

Our proposed algorithm uses the probability of detection accuracy values of attack intention to estimate the similar intentions. The probability values are conducted from (AIA) algorithm [7], which generates a list of probability attack intentions depends on relevant evidences. This combines Dempster–Shafer (D-S) evidence theory with causal networks to get a better estimation of the attack intentions.

3. Similarity of attack intentions algorithm

This section describes the similarity of attack intentions process model that estimates the similar intentions behind new attacks. The similarity metric that measures the attack intentions is generated to determine similar intentions.

The model is divided into three subcomponents, as shown in Fig. 1. The first subcomponent identifies the attack intentions of the current attack based on the same previous attack intention analysis process model through the new AIA algorithm. The similarity metric for attack intentions is generated in the second subcomponent to determine similar intentions. The third subcomponent utilizes the predefined attack evidence depository, which consists of previous intentions. The depository is utilized to select intentions based on the similarity between the intentions of the new attack and the previous ones. The components are described in the succeeding subsections.

3.1. Identification of attack Intention

This subcomponent adopts network capturing tools such as Wireshark, and Snort, as Network Intrusion Detection Systems (NIDSs). Network traffic is captured in the first step through network capturing tools, which normally produce a huge array of alerts and security data. A copy of the captured data is analyzed in the next step to identify the attack alerts. Based on these alerts, the model begins to collect evidence that relates to a specific attack. All the possible attack intentions are defined in the result.

3.2. Similarity metric generation for attack intentions

The attack intention probability is computed in this subcomponent with the AIA algorithm [7]; each value is associated with a relevant attack type to generate a similarity metric. The model employs a similarity metric to approximate the attack intention and to determine the similarity of the new attack intentions with the other pre-defined intentions.

3.3. Selecting similar intentions

This subcomponent selects the closest match to the intentions of the new attack. The model then utilizes a pre-defined database that contains data on all previous attacks and their real intentions. These data are collected during the investigation process.

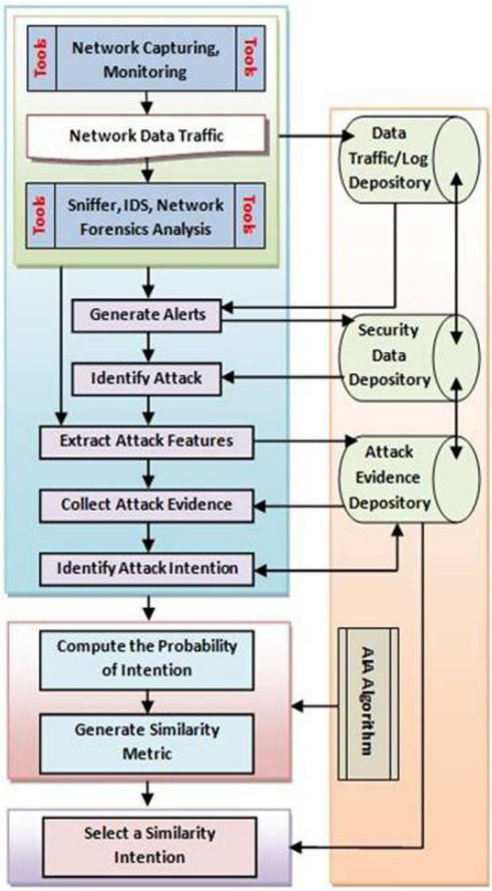

Fig. 1. Similarity of attack intentions process model based on (AIA) algorithm.

A new algorithm called Similarity of Attack Intentions (SAI) as shown in Fig. 2, which is based on the AIA algorithm and the proposed model, is designed to allow the subcomponents of the model to produce outcomes. This algorithm helps to better understand the logic of establishing the similarity of attack intentions process, which described in [13].

The SAI algorithm establishes the similar intentions between a new attack and previous ones. It defines a set named {PA}, which contains all previous attacks, where PA={A_1, A_2, A_3, … , A_n} and n is the integer number. Another set named {AI} is defined; it contains all attack intentions for all predefined attacks, where AI={I_1, I_2, I_3, … , I_n} and n is the integer number. One or more attack intentions from the set {AI} are relevant to one or more attacks from the predefined attack set {PA}, which means the relation between set {AI} and set {PA} is many-to-many.

A probability value is assigned to each attack intention in the set {AI} with the AIA algorithm [7]. We suppose that a new attack called A_k which belongs to the {PA} set occurs. This attack contains a set of intentions called A_kI_x, which is a subset of {AI} and $x \geqslant 1$. The SAI algorithm estimates the similarity values between the new attack intentions and the others. First, the algorithm identifies all the attacks that contain one or more of the attack intentions from the subset of {A_kI_x}. The algorithm computes the sum of all the probability values of the attack intentions that are relevant and similar to the A_k intentions for one attack. The similarity of attack Ak intention ($SimA_nI(A_k)$) is computed as the total probability value of the attack intentions (A_nI_x) divided by the total number of the similar intentions of a specific attack [13], as illustrated by the SAI algorithm.

The intentions similarity metric is generated from ($SimA_nI(A_k)$). We select the maximum similarity of attack intention value from the intentions similarity metric to identify the similarity between the intention of the new attack A_K and others; the value is computed by ($SimAI(A_k)$) [13], as illustrated by the SAI algorithm.

```
Input : Attack with their intentions probability.
Output: Estimate similar of the new attack intentions with
        other.
Begin
       Let PA a set of predefined attacks {A1,A2,A3…An}
       Define Ak as a new attack, where Ak ϵ PA
       Let AI a set of predefined attack intentions {i1,i2,i3…im}
       Define Ix using AIA as a set of all attack intentions for
              Ak,  where Ix ϵ AI
       Initialize the maximum similarity of attack intention
                  with Ak, MaxSimAI(Ak) = 0
       Initialize the summation of all similarity of attack
                  intention with Ak, SumSimAI(Ak) = 0
       For each An ϵ PA do
            For each Im ϵ AI do
                Select Id where Id ϵ Ix
                If Id founds, then
                  Assign AnId as the probability value of Id
                Else
                  Assign AnId = 0
                End If
               Compute SumSimAI(Ak) = SumSimAI(Ak) + AnId
           End For
       Compute SimAnI(Ak) = SumSimAI(Ak) / r, as a similar attack
                            intention, where r is the total number of Ak intentions
       If SimAnI(Ak) > MaxSimAI(Ak ), then
             Assign MaxSimAI(Ak) = SimAnI(Ak)
             Select n as an a maximum similar attack number
    End If
    End For
End
```

Fig. 2. Similarity of attack intentions (SAI) algorithm.

4. Implementation of SAI algorithm

This section introduces the SAI algorithm, which aims to implement the similarity of attack intentions process model based on an AIA algorithm. This algorithm presents a set of sequential operations to determine the new cyber crime intentions based on the predefined intentions.

The SAI algorithm establishes similar intentions. It depends on the AIA algorithm to assign the probability ratio for each intention. Similar intentions are established based on the similarity of the new attack intentions with the predefined intentions through the average value of the same intentions for each attack with the new one. For example, we assume that a new attack called A_n has three intentions: {i_1, i_2, i_3}. The SAI algorithm establishes a similar attack based on the attack intentions as shown in Table 1.

Table 1. Example of establishes a similar attack based on the attack intentions.

Attack ID	i_1	i_2	i_3	SumSim	SimA
A_1	0.23	0	0.15	0.38	0.126667
A_2	0.35	0.16	0.45	0.96	0.32
A_3	0	0	0.21	0.21	0.07
A_4	0.52	0.19	0.13	0.84	0.28
A_5	0.37	0.12	0	0.49	0.163333

Table 1 shows that the similarity of the new attack A_n with others is called attack A_2, which is the maximum similarity (SimA) value. The intention value (0) for A_1i_2, A_3i_1, A_3i_2, and A_5i_3 means that no same intentions among

these attacks are observed for attack A_n. SumSim presents the summation of the similar attack intentions for each attack, and SimA presents the average of the similar intention values.

5. Analyzing the experimental results

This section evaluates the SAI algorithm, which depends on the AIA algorithm to predict new cyber crime intentions and assign the probability values of accuracy detection to these intentions. A similar metric for new cyber crime intentions is generated to identify similar intentions.

We consider an example assumes that the new attack A_{11} intentions are analyzed with the AIA algorithm and detected as a set of four intentions: $\{i_1, i_4, i_6, i_2\}$. The potential intentions of the new cyber crime are $\{i_1, i_2\}$; the two have a maximum probability value and are reserved for the I_x set where x represents the number of intentions as shown in Table 2. The probability values for each intention present the accuracy of the height value of detection after the analysis process. Each value equals (0.68). We suppose that the predefined attack set is PA=$\{A_1, A_2, A_3, \ldots A_{10}\}$, and the predefined intention set is AI=$\{i_1, i_2, i_3, \ldots i_{10}\}$.

Table 2. Intentions probability values for all attacks.

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8	i_9	i_{10}
A_1	0.25	0	0.55	0	0.33	0	0.45	0	0	0
A_2	0.61	0.35	0	0	0	0	0	0.3	0	0
A_3	0.32	0	0	0	0	0.67	0.68	0	0	0.6
A_4	0.17	0.22	0.5	0.3	0	0	0	0	0	0
A_5	0.33	0.35	0	0	0.12	0.48	0	0	0	0.69
A_6	0.31	0	0.38	0	0.5	0	0	0	0.21	0
A_7	0.22	0	0.64	0.53	0	0.17	0.14	0.34	0	0
A_8	0.35	0.21	0	0	0	0	0.29	0	0.35	0
A_9	0	0.62	0.43	0	0	0	0	0	0	0.2
A_{10}	0.02	0.56	0.47	0	0.7	0	0.68	0.13	0.19	0
A_{11}	0.68	0.68	0	0.17	0	0.43	0	0	0	0

The SAI algorithm identifies the similar intentions $\{i_1, i_2\}$ between the new attack A_{11} and previous ones as shown in Table 3.

Table 3. The similar intention values for i_1 and i_2.

	Similarity with (i_1)	**Similarity with (i_2)**	**Similarity with (i_1) & (i_2)**
A_1	0.25	0	0.125
A_2	0.61	0.35	0.48
A_3	0.32	0	0.16
A_4	0.17	0.22	0.195
A_5	0.33	0.35	0.34
A_6	0.31	0	0.155
A_7	0.22	0	0.11
A_8	0.35	0.21	0.28
A_9	0	0.62	0.31
A_{10}	0.02	0.56	0.29

Fig. 3 shows the similarity values between each intention in attack A_{11} and the predefined attacks.

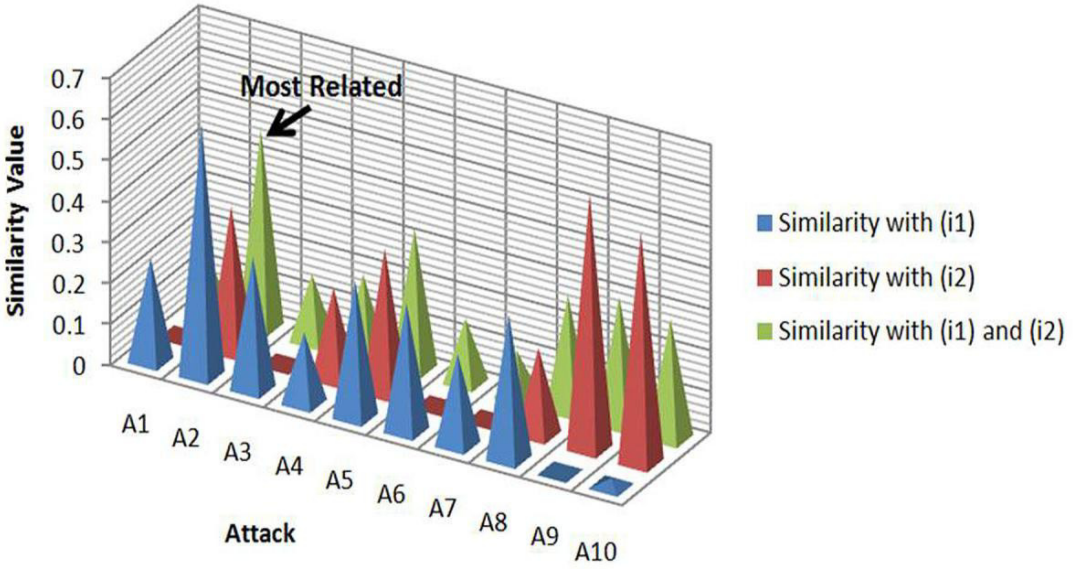

Fig. 3. Similarity relation between the intentions of A_{11} with other attacks.

The evaluation of the results shows that the SAI algorithm provides useful information and increases the possibility value of evidence in advance, thereby helping investigators eliminate similar cases based on the similarity of intentions. The results present the relation between the new cyber crime and other crimes, which helps improve the decision making process.

6. Conclusion and future work

In this paper we proposed a new algorithm called Similarity of Attack Intention (SAI), to estimate the similar cyber crimes intentions. Based on (AIA) algorithm we identified the probability values of accuracy detection for the cyber crimes intentions. The results explained there is a relation between the new cyber crime and the pre-defined cyber crimes. It also, proved that the similarity measurements of the intentions help the investigators to estimate the similar cases of the new cyber crime with others, in order to reduce the time and processing cost.

Attack intentions can be used as an efficient factor to identify the attack strategy in advance to increase the possibility value of evidence in network forensics. Moreover, the accuracy of the estimated values of the similar cyber crime could be increased when the Case-Based Reasoning technique is used.

Acknowledgements

This work was supported by MOSTI ScienceFund Grant No. 305/PKOMP/613144, School of Computer Science, Universiti Sains Malaysia, Penang, Malaysia.

References

[1] CERT, CSO, and U.S.S. Service, Cyber Security Watch Survey. Software Engineering Institute CERT Program at Carnegie Mellon University and Deloitte. 2011.

[2] Palmer, G., A Road Map for Digital Forensic Research, in Report from DFRWS, F.D.F.R. Workshop, Editor.: Utica, New York. 2001: p. 27–30.
[3] Shin, Y.-D., New Model for Cyber Crime Investigation Procedure. JNIT: Journal of Next Generation Information Technology, 2011. 2(2): p. 1-7.
[4] Pilli, E.S., R.C. Joshi, and R. Niyogi, Network forensic frameworks: Survey and research challenges. Digital Investigation, 2010. 7(1-2): p. 14-27.
[5] Almulhem, A. Network forensics: Notions and challenges. in Signal Processing and Information Technology (ISSPIT), 2009 IEEE International Symposium on. 2009.
[6] Huang, M.-Y., R.J. Jasper, and T.M. Wicks, A large scale distributed intrusion detection framework based on attack strategy analysis. Computer Networks, 1999. 31(23-24): p. 2465-2475.
[7] Rasmi, M. and A. Jantan, AIA: Attack intention analysis algorithm based on D-S theory with causal technique for network forensics - A case study. International Journal of Digital Content Technology and its Applications, 2011. 5(9): p. 230-237.
[8] Rasmi, M., et al., Attack Intention Analysis Model for Network Forensics, Software Engineering and Computer Systems. Springer Berlin Heidelberg. 2011. p. 403-411.
[9] Wu, P., W. Zhigang, and C. Junhua. Research on Attack Intention Recognition Based on Graphical Model. in Information Assurance and Security, 2009. IAS '09. Fifth International Conference on. 2009.
[10]Peng, W., S. Yao, and J. Chen. Recognizing Intrusive Intention and Assessing Threat Based on Attack Path Analysis. in Multimedia Information Networking and Security, 2009. MINES '09. International Conference on. 2009.
[11] Wei, W. and E.D. Thomas, A Graph Based Approach Toward Network Forensics Analysis. ACM Trans. Inf. Syst. Secur., 2008. 12(1): p. 1-33.
[12] Zaka, B., Theory and Applications of Similarity Detection Techniques, in Institute for Information Systems and Computer Media (IICM)., Graz University of Technology: Graz, Austria. 2009.p. 171.
[13] Jantan, A., et al., A Similarity Model to Estimate Attack Strategy Based on Intentions Analysis for Network Forensics, in Recent Trends in Computer Networks and Distributed Systems Security., Springer Berlin Heidelberg. 2012. p. 336-346.